

AI+RTC：如何用「实时音视频技术」 创造「AI实时对话自然交互」

目录 Menu

- AI对话应用场景
- TRTC实时对话式AI解决方案
- TRTC产品介绍和核心优势

前言

Voice Agent 可能是今年最值得关注的一个方向，是下一代人机交互界面，多模态改变交互

GUI → LUI

语音交互以其与人类自然沟通方式的高度一致性，提供了一种更为直观和低能耗的交互体验，而打字文本交互的使用门槛更高

低延迟、高智能

GPT-4o 是第一个实现端到端 voice-in, voice-out 的大模型，低延迟、真实情感表达、很强的语义理解能力

重新发明电话

围绕实时对话构建的产品受到投资人青睐，利用AI定制特定类型的实时对话，更高效地提供上下文和价值

AI对话应用场景

自24年下半年以来，涌现了哪些新场景？

01

大模型/搜索：实时AI对话已经成为标配

05

AI+游戏：AI陪玩、AI NPC、AI队友

02

AI+泛娱乐：AI陪伴、语聊房助手

06

AI+健康：问诊助手、心理健康咨询

03

AI+教育：实时字幕、在线课堂助手

07

AI+客服：智能客服、智能外呼

04

AI+IoT：可穿戴设备、AI玩具

08

AI+工作：招聘、实时翻译

应用场景01：大模型/搜索——实时AI对话已经成为标配

大模型厂商



从「会思考」到「能对话」

RTC+AI的方案不断优化

- 在嘈杂的地铁场景中，用户仍可通过语音与大模型完成复杂问题交互，且支持随时打断和话题切换，接近真人对话体验

有效提升对话轮次和客户粘性

- 相比仅用IM单次对话，加入AI实时语音对话平均轮次提升3-5倍



未来挑战与演进方向

长对话意图漂移

- 当前仍需解决长时对话中的意图漂移问题（如30轮以上对话的上下文丢失率约12%）

多语种混合准确性

- 多语种混合场景的实时翻译精度（如中英文混杂语句的识别错误率约8%）

应用场景02: AI+泛娱乐沿着虚拟陪伴和互动内容两条路进化

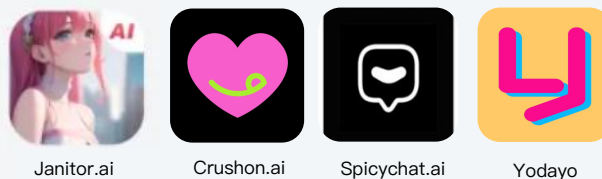
AI陪伴分类

情感陪伴



NSFW

(Not Safe For Work)



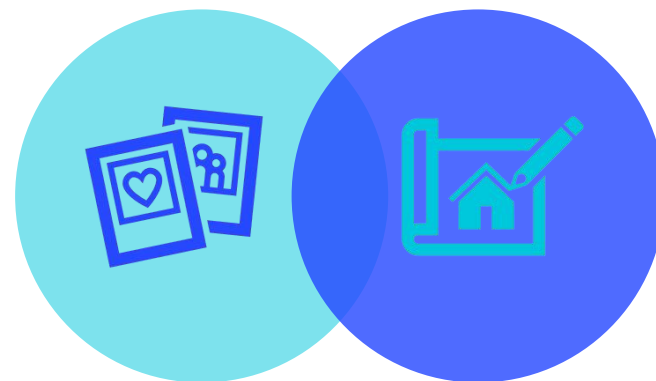
- 虚拟体验需求，一段虚拟的剧情体验，用户使用该产品的目的性比较明确，**通过和AI的互动获得「情绪价值」**
- 想办法满足更广泛人群的虚拟体验需求，**成为一个大DAU的产品**
- NSFW可能是目前AIGC产品里**停留时长和付费意愿最强的品类**
- 据了解有些做海外的已经实现了盈利（主要靠擦边和强收费）

目前阶段的 AI 虚拟陪伴，**还有很多难点需要解决**

- 1.记忆力不足
- 2.成本较高
- 3.没办法完美模仿一个 IP
- 4.文字太单薄，需要多模态

虚拟陪伴

互动内容



今天的 C.AI，并不是真正的恋爱陪伴，而是一种新的“内容消费”形式，是一种全新的互动内容。



用户和 AI 在共创一个故事，并且不同于看小说，AI 每次回复的内容，可能都是不一样的，会往完全不同的剧情发展。

应用场景03：对话式AI在教育行业的应用场景

- LLM+Voice Agent是下一代交互的开始：低延迟、高智能。
- 长链路的延迟是用户难以接受的，RTC+LLM使得延迟降低，加入了智能打断等功能让用户交互起来更像在于真人交流。
- 这对于互动/实时要求性高的教育场景具有显著意义。典型场景如语言学习、儿童兴趣引导、阅读伴侣等。

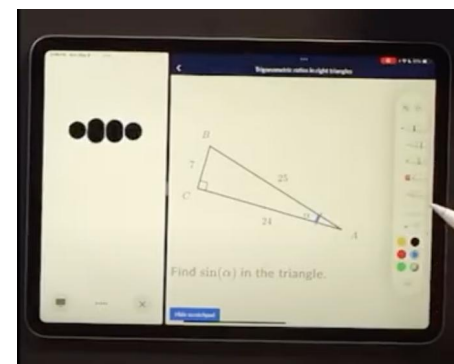
实时字幕+AI助手同声传译



语言学习、口语陪练



多模态能力解题讲解



AI大模型参与教学的能力逐渐增强，难度逐渐增大

软件类

硬件类



儿童陪伴：实时阅读伴侣



单词笔：语言学习、实时对话



AI学习机：实时辅导

应用场景04: IoT AI对话

(一) 手机&电脑硬件行业

智能手机

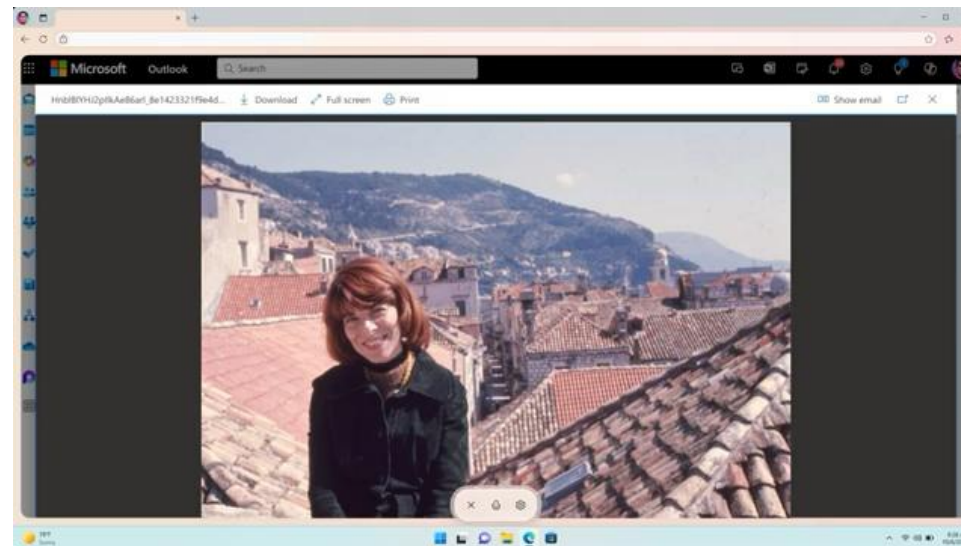


Google Pixel 9 with Gemini Live

**搭载了大模型的手机语音助手
响应更快，质量更高**

国产手机厂商正在加快脚步进行调研
这里是我们的机会点

电脑助手



**搭载了大模型的电脑语音助手
可以与 AI 助手聊天、向它提问，并像与朋友或
同事交谈一样打断它**

应用场景04: IoT AI对话

(二) 智能家居行业

扫地机器人



科沃斯 | YIKO语音2.0大语言模型技术

更好的自然语言理解，清晰指令

追觅、科沃斯、石头、云鲸等头部玩家均在比拼

智能摄像头

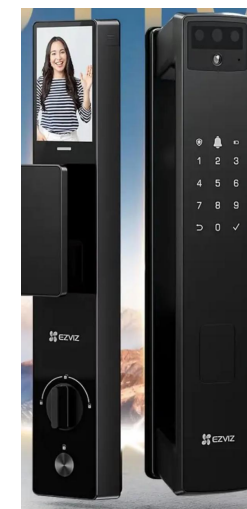


Enabot 赋之 Ebo Air 智能机器人

陪伴机器人

从摄像头延伸出来的宝宝陪伴、陪伴机器人都有机会

智能门铃



萤石AI三摄全自动人脸视频锁
行业首款接入蓝海大模型的智能锁

通过摄像头实时识别并理解视频内容

多模态视频能力落地的典型场景

应用场景04: IoT AI对话

(三) 智能穿戴设备

智能手表



儿童手表

智能手表的屏幕较小，对语音交互的诉求更强

小天才、360手表、其他白牌厂商均有机会

智能眼镜

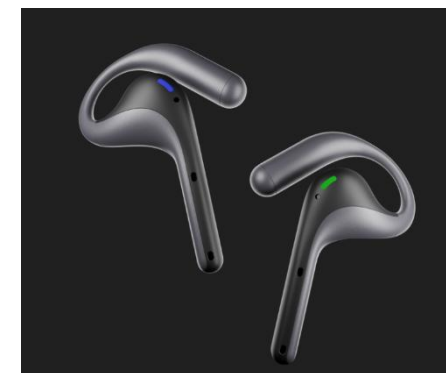


Ray-Ban Meta智能眼镜

**在智能眼镜中加入摄像头、AI 等功能
通过 AI 提升语音交互能力**

智能眼镜可能是AI+多模态最适合的场景，智能眼镜是最完美的AI助手，行业还在早期，未来还会有新的玩家出来，密切关注

智能耳机



时空壶W4Pro

智能耳机目前主要应用在1对1翻译，支持双方对话过程随时发言

创业公司+老牌耳机公司均有机会

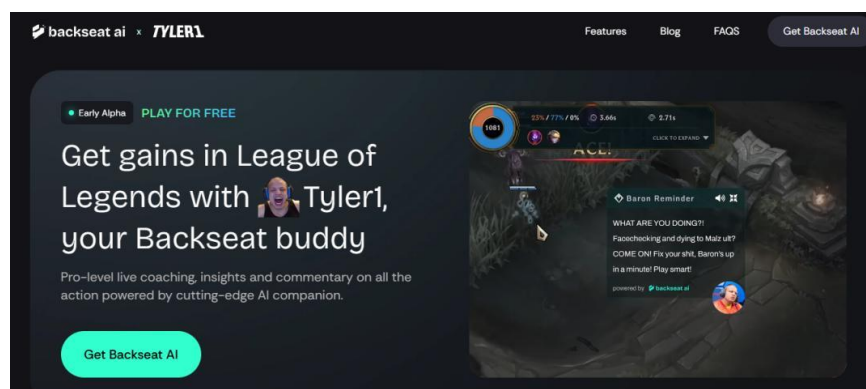
应用场景05: 对话式AI在游戏中的应用场景

大模型参与游戏流程的能力逐渐增强

1. AI NPC



2. AI陪玩/指导



3. AI队友: 接收指令/执行



应用场景06: 对话式AI在医疗健康行业的应用



智能辅助诊疗与决策支持

- AI诊疗助手
- 疑难病例多学科会诊
- 个性化药物推荐



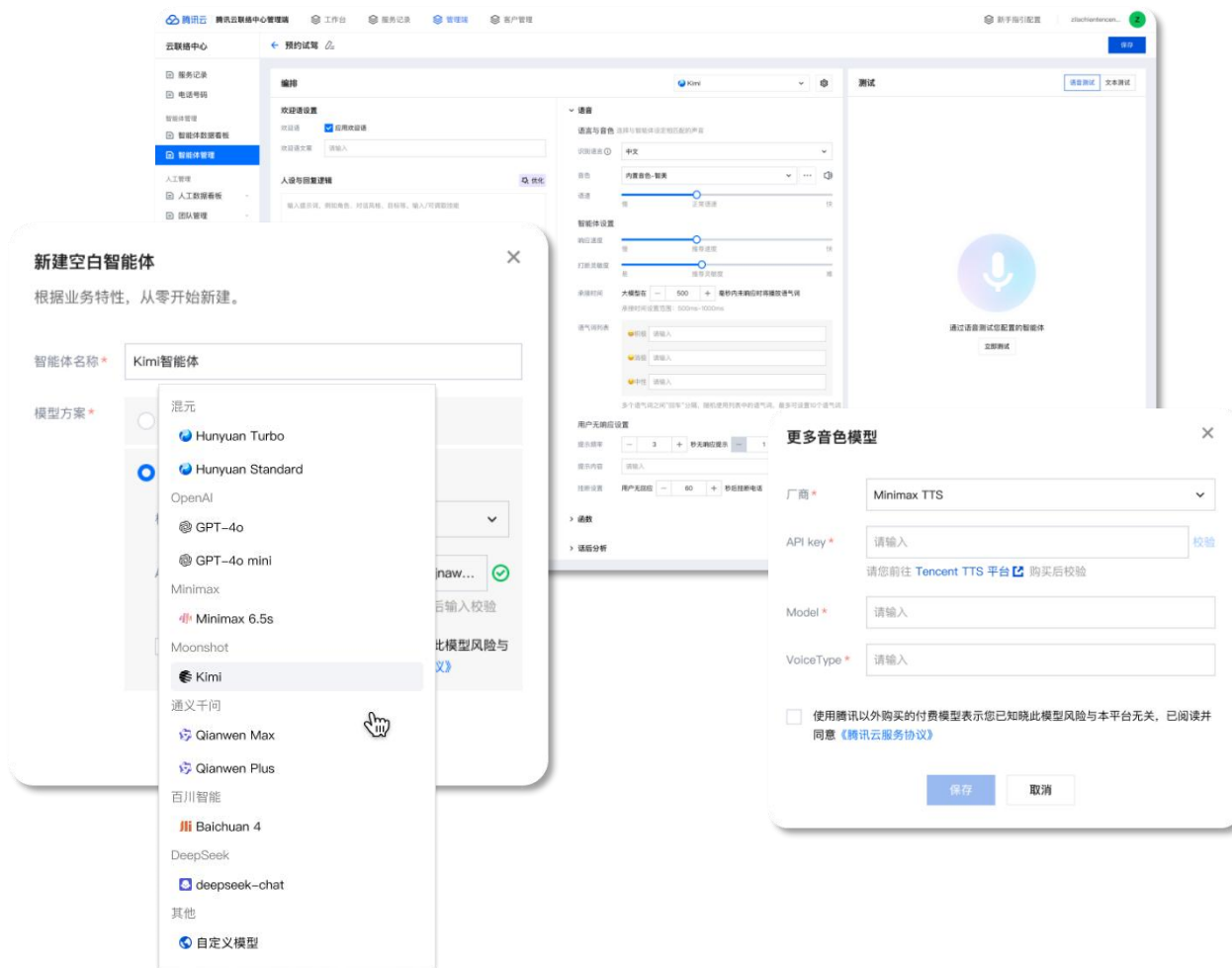
心理健康支持与危机干预

- 24小时情感疏导与心理评估
- 虚拟现实疗法与社交训练

应用场景07: AI+客服——快速实现智能外呼和智能呼入能力

配置智能体

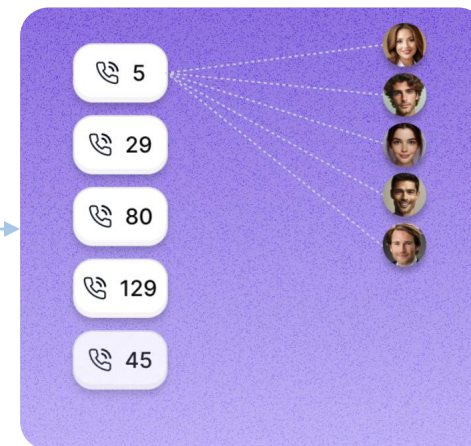
创建语音智能体，开放接入大模型，支持接入自定义TTS音色并配置效果。



发起外呼



调用外呼接口



实现呼入

绑定智能体至电话号码，用户呼入后由智能体接待。

The form shows the configuration for incoming calls. It includes a '电话号码' (Phone Number) field with the value '00860755...492'. Below it is a '呼入设置' (Incoming Call Settings) section with a '选择智能体' (Select AI Agent) dropdown menu, which is currently set to 'Kimi智能体'.



拨打400或固话号码



2025趋势：多模态交互升级、垂直场景深化、物理实体走得更远

关注点

潜在机会

多模态交互升级

- 语音交互更加自然流畅，能够准确识别各种口音和复杂语义，用户通过语义打断
- 视觉交互取得新突破

视频多模态的应用场景如何和AI陪伴结合起来？

垂直场景深化

- 垂直场景进一步深化，目前还属于探索期，25年有可能迎来新的爆发，例如，教育领域出现单月使用超100小时的AI辅助教学场景

声音的 agent 和泛娱乐各个场景的结合是不是能孕育一个新交互模式的抖音应用机会？

物理实体

- “物理实体”是“陪伴赛道”中与“智能”同等重要的要素：搭配了物理实体的“AI陪伴”产品可以走得更远

消费电子终端Agent化能够带来哪些新的应用场景和机会？

TRTC实时对话式AI 解决方案

方案概述

现有的AI语聊场景主要是基于IM场景的离线文字聊天或者语音聊天，GPT-4o的发布，把多模态大模型的应用场景提升到了实时的语音或者视频交互。

使用Tencent RTC，结合第三方的大模型和TTS等，您可以轻松创建一个媲美GPT-4o所演示的AI实时交互体验，应用到您的业务场景内。



功能特点和技术优势

功能特点

实现超低延迟的AI对话

音视频端到端延迟低于**300ms**，AI对话延迟低于**1000ms**。

精准的ASR识别

支持多种语言，包括英语、西班牙语、日语、韩语、中文以及**23种方言**和**130种国际语言**。

第三方LLM和TTS无缝集成

配置LLM和TTS服务的账户凭证，**无缝集成**到服务后台。

高兼容性

支持多个平台，兼容超过**20,000**种设备模型。



红框内客户自定义

- **腾讯侧：**客户端和服务端的一体化集成方案，提供智能打断、降噪能力。
- **客户侧：**自定义大模型和TTS，接入到腾讯AI服务，保持业务灵活性。

技术优势

双讲打断

AI聊天过程，实时打断，真正实现与人聊天的真实体验。

AI降噪

全新降噪引擎，输入端音频实时AI降噪，提升ASR的识别准确度。

服务稳定

全球2800+节点，多服务容灾，高稳定超低延迟。

DEMO

超低延迟

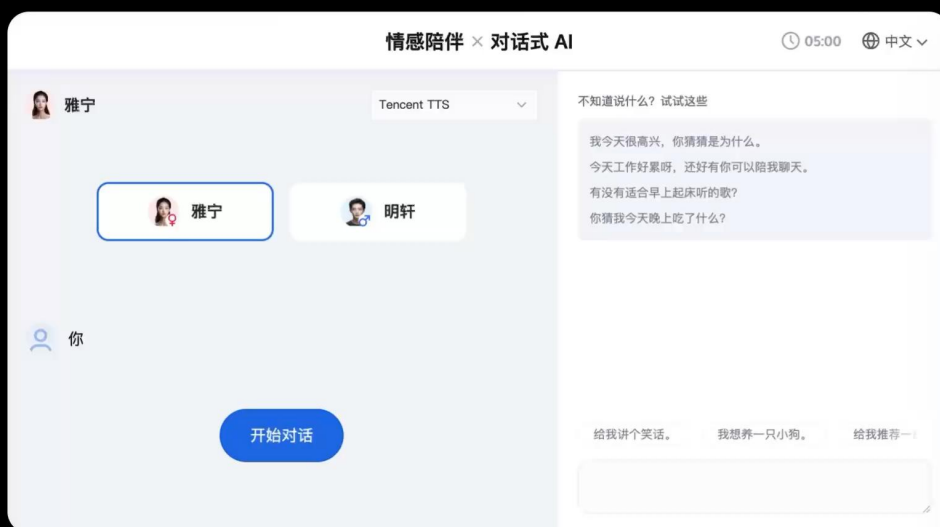
对话总延迟低至1000ms以内

智能打断

丰富的打断方式，提升对话自然度

上下文管理

强大的上下文记忆，对话更流畅



WEB DEMO在线体验：
<https://rtcube.cloud.tencent.com/prerelease/component/experience-center/index.html#/detail?scene=ai>

延迟优化：通过预测、缓存和智能分片等技术，提升响应速度

大模型预请求

- ✓ 实现对话意图预测，通过适当浪费大模型请求来降低大模型延迟

动态插入语气词/寒暄语

- ✓ 设计语气词库，实现上下文相关的寒暄语生成，大幅降低体感延迟

基于语义的句子边界识别

- ✓ 实现智能分片边界判断，精准识别网址、markdown格式、公式等复杂文字的边界判断，适配更多语种

List of the processing steps in the voice-response loop

Project	Latency (ms)
Mic input	40ms
Opus encoding	21ms
Network stacks and transit	10ms
Pack handling	2ms
Jitter buffer	40ms
Opus decoding	1ms
STT	300ms
LLM TTFB	300ms
TTS TTFB	200ms
Opus encoding	21ms
Packet handling	2ms
Network stacks and transit	10ms
Jitter buffer	40ms
Opus decoding	1ms
Speaker output	15ms
Total ms	1003ms

全流程流式

- STT/LLM/TTS全流程流式处理，减少延迟

网络优化

- 内部使用内网传输，外部供应商支持GBN加速

欢迎语优化

- 支持欢迎语缓存，减少TTS费用和延迟

自部署方案

- 采用自部署STT/LLM/TTS方案，进一步减少延迟

调整播放buffer

- AI对话场景码率更低，配置更低的播放buffer，减少延迟

语音识别与降噪优化方案

识别准确度问题

- 专业术语与行业专有名词识别困难
- 口语化表达场景下字错率偏高
- 语音3A处理影响识别精度



优化策略

场景化优化

- 为不同客户定制专属热词词库
- 基于真实场景音频数据进行标注与模型微调

降噪增强

- 针对性开发远场声音消除功能，提升AI通话质量
- 服务端AI降噪支持，支持小程序/Web/电话等多端场景

联合优化

- 降噪与ASR团队协作
- 通过A/B测试，优选最佳参数组合

降噪优化：提升语音识别质量

远场人声消除

- 应用深度学习模型，实现远场人声的精确识别和消除

说话人声分离

- 使用盲源分离算法，再多人对话场景中提取目标说话人的声音

声纹识别

- 建立声纹特征提取模型，实现实时说话人身份确认和切换检测

LLM 技术优化方案

极致低延迟

- 大模型访问默认流式传输
- 引入连接池机制，降低建联延迟
- 对大模型输出进行流式分片，保证最小的延迟

对话能力增强

- 内置上下文管理系统
- 支持自然流畅的多轮对话

生态兼容性

- 支持兼容 OpenAI API 规范的厂商
- 无缝对接 Coze/Dify 等主流 Agent 平台
- 支持 RAG 知识库增强能力，支持 Function Call 功能扩展

TTS部分技术优化方案

流式处理

TTS全部采用流式，保证最低延迟

多平台支持

已支持TTS厂商包括

- 腾讯
- Minimax
- Azure
- 火山
- Elevenlabs
- Cartesia

允许自定义TTS接入

分片优化

配合大模型流式分片引擎，保证低延迟和分片正确性，支持客户级别的分片规则下发

智能打断：分为自动打断和手动打断，有效提升对话的流畅性和自然度

自动打断



手动打断

自动打断是指系统在检测到特定条件时，根据用户的音频输入自动触发，支持两种主要规则：基于**音频时长的打断**和**基于语义的打断**。

1) 基于音频时长的打断

- 用户可动态配置打断时长，默认为500ms
- 可调范围：300ms（更灵敏）到5000ms（避免误打断）
- 建议：
 - 对于客服对话等高交互场景，设置较短时长(如300ms 或者 500ms)以提高响应速度
 - 对于演讲或授课场景，设置较长时长(如1000ms)以避免误打断

2) 基于语义的打断

基于语音活动检测（VAD）技术，当服务端检测到用户输入了一句语义完整的话时，会自动触发打断。

手动打断是指由用户主动发起，通过端上SDK发送自定义消息实现，为用户提供更直接的控制方式。通常依赖于用户界面上的按钮或快捷键。

1) 打断信令说明

使用JSON格式的自定义消息

2) 使用方法

通过SDK的sendCustomCmdMessage方法发送上述自定义消息即可实现手动打断。

打断优化：提升交互自然度

01

打断准确性提升

优化VAD模型，降低背景噪声触发打断的概率；增加语义打断能力，结合语义判断是否要打断

03

打断延迟优化

优化实时性能

02

打断自然度提升

避免出现抢话的场景，优化打断时的语音合成效果

04

ASR误识别修正

ASR错误检测，基于大模型对误识别进行修正

TRTC产品介绍和核心优势

Tencent RTC：快速实现全球跨平台、超高品质的实时音视频互动场景

超高品质

高品质：腾讯会议同款3A引擎

高抗性：丢包率80%正常语音通话

低延时：全球平均端到端延时<300ms

低门槛快速接入

提供覆盖泛娱乐、教育、企业服务丰富的场景化组件以及全平台的代码示例

官方支持Flutter、Electron跨平台框架

跨平台全球互通

端：Android、iOS、Windows、Mac、Web全平台SDK。

云：覆盖全球200+国家和地区的实时传输网络

高质量全球部署

提供包括亚太、北美、欧洲、中东、非洲、

拉丁美洲等覆盖全球的海外接入点与加速点

基础功能

超低延时直播场景

<1s 延时观看



多人实时互动场景

<300ms 实时互动



30亿

每日支撑上行时长

5000+

每月使用客户数量

合规及隐私

通过ISO/CSA/NIST/BS 7项合规认证
并符合GDPR、CCPA规范

产品矩阵



Tencent RTC-数据安全与稳定

20,000₊终端设备

多平台互通，支持 iOS, Android, Windows, macOS, web, Flutter, Electron, Unity, Unreal等，20,000+终端设备完美适配



200₊国家地区

全球节点布局，丰富的高带宽资源储备，跨越200+国家地区，提供覆盖全球的高连通、高可靠、强安全的网络连接通道。



<300_{ms} 延迟

自研多重最优寻址算法，具有全网调度能力，端到端平均时延 < 300ms，在丢包率70%、网络抖动1000ms以上的情况下，仍能流畅进行视频通话



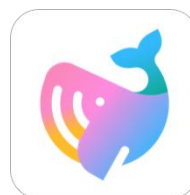
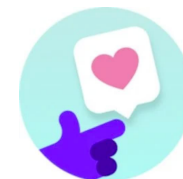
Tencent RTC — 全球覆盖

亚洲	中国	全覆盖(含香港、台湾)
	印度	孟买
		印度南-金奈
		印度南-金奈
		印度东-加尔各答
		印度北-新德里
	新加坡	新加坡
	泰国	曼谷
	韩国	首尔
	日本	东京
	菲律宾	马尼拉
	马来西亚	吉隆坡
	越南	胡志明
		河内
	印度尼西亚	雅加达
	蒙古	蒙古
	柬埔寨	柬埔寨

中东	阿联酋	阿联酋
	沙特	沙特
	迪拜	迪拜
	科威特	科威特
	卡塔尔	卡塔尔
	阿曼	阿曼
欧洲	德国	法兰克福
	英国	伦敦
	土耳其	伊斯坦布尔
	葡萄牙	里斯本
	西班牙	马德里
	意大利	意大利
南北美	美国	圣克拉拉
		阿什本
		迈阿密
		蒙特利尔
	墨西哥	墨西哥
	巴西	巴西
	哥伦比亚	波哥大

非洲	南非	南非
	埃及	埃及
	尼日利亚	尼日利亚
大洋洲	澳大利亚	墨尔本
	澳大利亚	悉尼
	新西兰	新西兰

典型合作客户



感谢观看！

Thank you

AI语音交互软硬件的实践路径与 场景创新

魔镜云科技董事长：曾 随 城

2025年4月17日



目 录

CONTENTS

1

第一章 公司简介与产品定位

2

第二章 项目里程碑和开发经验

3

第三章 产品案例与应用场景

4

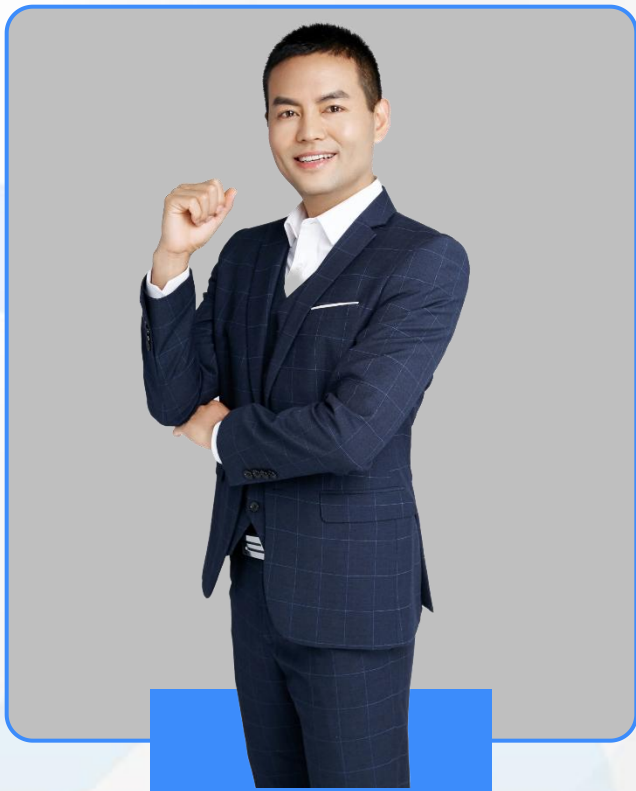
第四章 市场趋势与未来方向

5

第五章 结束语与合作展望

讲师介绍

咨询导师：曾 随 城



人工智能训练师【三级/高级】

国家一级人力资源师

深圳市先进教育工作者

ACI注册国际职业培训师

提供 AI 技能培训、AIGC工具研发、AI软件设计、
AI 模型建立，企事业单位 AI 战略咨询。

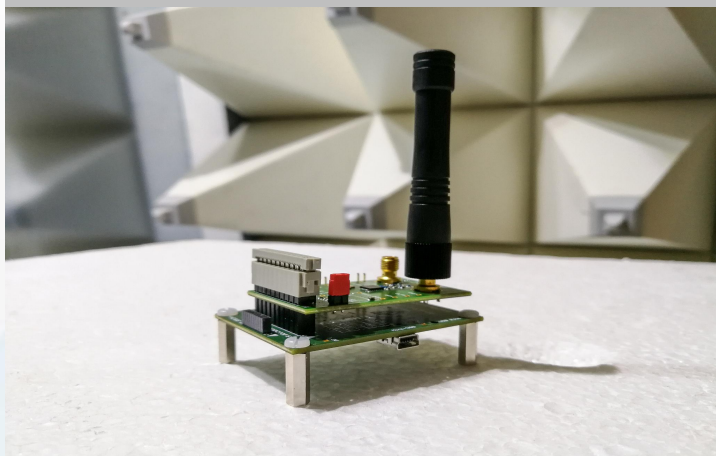
目录 Menu

- 公司简介与产品定位
- 项目里程碑和开发经验
- 产品案例与应用场景
- 市场趋势与未来方向
- 结束语与合作展望

魔镜云科技简介

公司背景

魔镜云科技成立于2019年，专注于AI语音交互硬件的研发与创新，致力于通过智能技术提升用户的生活质量和工作效率。



市场定位

魔镜云科技的产品主要面向智能家居、健康管理、企业服务等领域，旨在为用户提供高效、便捷的智能解决方案。

技术实力

- 依托东华大学、华中科技大学、华中师范大学、深圳大学等著名研究机构，专注于AI应用开发、模型训练。
- 出版书籍被选为大学虚拟现实专业的教材





➤ 我们是谁？

专注AI语音交互软硬件一体解决方案

- ✓ 聚焦“语音交互 + 场景AI模型”的终端产品开发
- ✓ 用软硬件一体化方案，解决失眠、情绪、销售管理等真实痛点
- ✓ 已推出：AI哄睡师、AI心理大师、AI销售助理

目录 Menu

- 公司简介与产品定位
- **项目里程碑和开发经验**
- 产品案例与应用场景
- 市场趋势与未来方向
- 结束语与合作展望

关键里程碑

项目节奏清晰，快速推进：



10月

完成语音技术方案选型



11月

基于TRTC完成Web Demo，赢得客户信任



12月

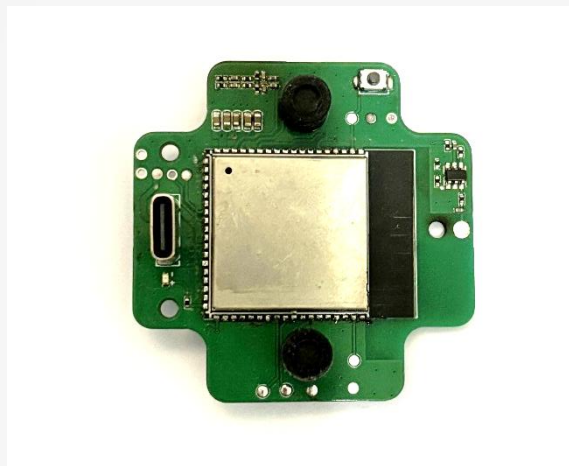
开发首批语音硬件原型

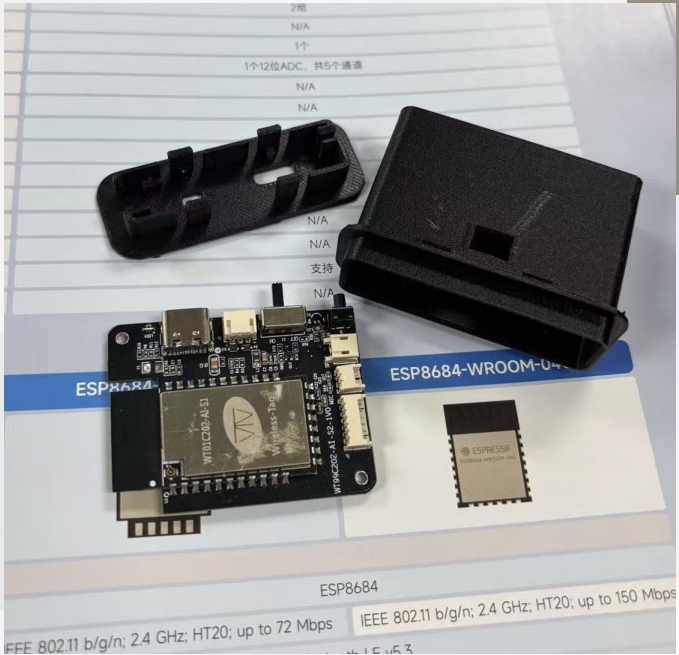
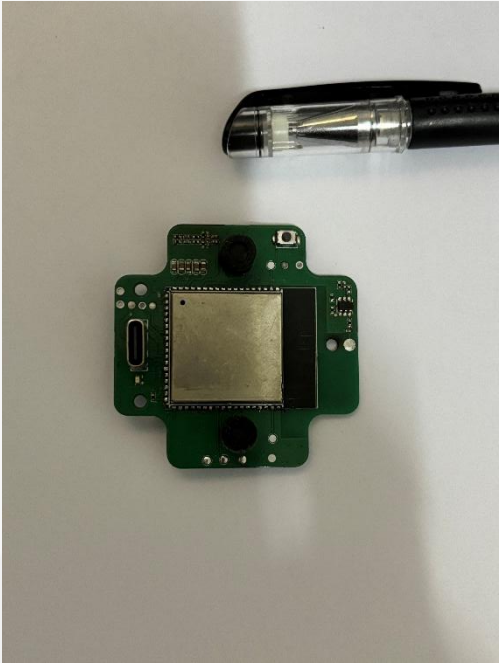


1~3月

3轮版本优化 + UI交互升级 + 语音细节调优

结果：推动多个合作订单签署





目录 Menu

- 公司简介与产品定位
- 项目里程碑和开发经验
- **产品案例与应用场景**
- 市场趋势与未来方向
- 结束语与合作展望

产品案例一：AI哄睡师

AI哄睡师通过自然语言处理技术，能够与用户进行睡前对话，提供舒缓的语音互动，帮助用户放松心情，缓解焦虑，为入睡创造良好的心理环境。

个性化音乐推荐

基于用户的睡眠习惯和偏好，AI哄睡师能够推荐适合的助眠音乐或白噪音，并通过语音控制播放，帮助用户更快进入睡眠状态。

智能语音互动

AI哄睡师能够模拟人类的语气和情感，提供温暖的睡前陪伴，尤其适合独居老人或儿童，帮助他们缓解孤独感。



睡眠数据分析

AI哄睡师能够通过内置传感器或与智能设备联动，监测用户的睡眠质量，生成详细的睡眠报告，并提供改善睡眠的建议。

情感陪伴

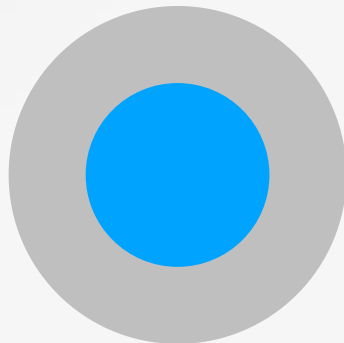
定位：帮助睡眠的智能语音陪伴者

产品案例二：AI心理大师



情绪识别与干预

AI心理大师通过语音识别和情感分析技术，能够实时监测用户的情绪状态，并提供针对性的心理疏导建议，帮助用户缓解压力、焦虑或抑郁情绪。



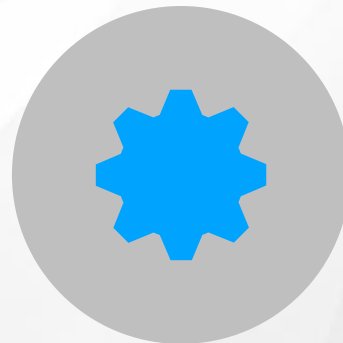
心理测评与建议

AI心理大师内置专业的心理测评工具，能够通过语音交互完成用户的心理健康评估，并根据结果提供个性化的心理支持方案。



情感陪伴与倾听

AI心理大师能够模拟心理咨询师的角色，提供无偏见的情感倾听服务，帮助用户表达内心感受，缓解心理负担。



危机干预与资源推荐

在用户出现严重心理问题时，AI心理大师能够识别危机信号，并提供紧急干预措施，同时推荐专业心理咨询资源或热线服务。

定位：轻心理问题场景的智能陪伴式支持

产品案例三：AI销售助理

智能客户分析

AI销售助理能够通过语音交互收集客户需求，并利用大数据分析技术，为客户提供个性化的产品推荐，提升销售转化率。

实时数据反馈

AI销售助理能够实时监控销售数据，并通过语音播报或可视化报告的形式，为销售团队提供业绩分析和市场趋势洞察，帮助优化销售策略。

销售流程自动化

AI销售助理能够自动完成客户跟进、订单处理、售后服务等环节，减少人工操作，提高销售效率并降低错误率。

多语言支持与跨区域管理

AI销售助理支持多语言交互，能够协助跨国企业或跨区域团队进行销售管理，打破语言和地域限制，提升全球化运营效率。

定位：智能化的语音渠道运营助手

目录 Menu

- 公司简介与产品定位
- 项目里程碑和开发经验
- 产品案例与应用场景
- **市场趋势与未来方向**
- 结束语与合作展望

市场趋势分析

智能家居需求增长

随着智能家居市场的快速发展，AI语音交互硬件作为核心控制设备，其需求持续增长。消费者希望通过语音指令实现家电控制、安防监控等功能，推动了相关产品的普及。



个性化交互体验

用户对语音交互的个性化需求日益增强，AI语音硬件正在从简单的语音识别向情感识别、语境理解等方向发展，以提供更加自然和贴心的交互体验。



跨设备互联

AI语音硬件正逐渐成为跨设备互联的枢纽，通过与智能手机、智能手表、车载系统等设备的无缝连接，构建更加完整的智能生态系统。



隐私与安全关注

随着AI语音硬件的广泛应用，用户对数据隐私和设备安全的关注度显著提升。厂商需加强数据加密和隐私保护技术，以赢得用户信任。



未来发展方向



多模态交互融合

未来的AI语音硬件将结合视觉、触觉等多种交互方式，提供更加丰富的用户体验。例如，通过摄像头捕捉用户表情和手势，结合语音指令实现更精准的交互。



边缘计算与本地化处理

为提升响应速度和保护用户隐私，AI语音硬件将更多地采用边缘计算技术，在本地完成数据处理和分析，减少对云端的依赖。



行业定制化解决方案

针对不同行业的需求，AI语音硬件将提供定制化的解决方案。例如，在医疗领域，语音硬件可用于病历记录和患者咨询；在教育领域，可用于语言学习和课堂互动。



情感计算与人性化设计

未来的AI语音硬件将更加注重情感计算能力，能够识别用户情绪并作出相应反馈，同时在外观设计上更加注重人性化和美观性，融入家居环境。

目录 Menu

- 公司简介与产品定位
- 项目里程碑和开发经验
- 产品案例与应用场景
- 市场趋势与未来方向
- **结束语与合作展望**

合作邀请：未来已来 探索无限

生态共建 技术共享

魔镜云科技通过持续的技术创新和产品迭代，与合作伙伴共同构建开放的AI语音交互生态，涵盖硬件、软件、服务等多个维度，通过跨界合作与资源整合，打造覆盖教育、办公、家庭等多场景的智能化解决方案，为用户提供无缝衔接的智能体验。





感谢观看！

Thank you